General Hand-Eye Calibration Based on Reprojection Error Minimization

Kenji Koide[1] and Emanuele Menegatti[1]

***Abstract*—This paper describes a novel hand-eye calibration technique based on reprojection error minimization. In contrast to traditional hand-eye calibration methods, the proposed method directly takes images of the calibration pattern and does not require to explicitly estimate the camera pose for each input image. The proposed method is implemented as a pose graph optimization problem, so that it can solve the estimation problem efficiently and robustly, and it can be easily extended for different projection models. It can deal with different camera models (e.g, X-ray cameras with a source-detector projection model) by changing the projection model. Through simulations, we validated that the proposed method shows a good estimation accuracy, and it can be applied to hand-eye calibration with a source-detector camera model. The experimental results with real robots show that the proposed method is applicable to real environments, and it improves the quality of a task which requires accurate hand-eye estimation, like 3D reconstruction.**

***Index Terms*—Calibration and Identification, Industrial Robots**

I. INTRODUCTION

HAND-EYE calibration is the task of estimating the spatial transformation between the end effector frame of a robot hand and the optical frame of a camera mounted on the end effector. It is essential for tasks where the robot moves the camera to a specific position and maps visual measurements into the robot frame (e.g., product inspection [1] and object picking [2]). It has been widely studied in the robotics and the computer vision communities, and a number of hand-eye calibration methods for RGB and depth cameras have been proposed.

In the context of SPIRIT[2], an industrial inspection robot project, an accurate and "general" hand-eye calibration method is required. This project aims to develop an automatic and general visual inspection robot framework, in which the camera for inspection can be easily replaced with any kind of imaging sensor suitable for any new inspection task, to name a few: RGB-D, thermographic, and X-ray cameras. The challenge here is to deal with those cameras with different characteristics in a unified framework. For instance, industrial thermographic cameras have sometimes a very narrow field of view, while the working distance of structured-light depth cameras is very limited. In particular, the optics of X-ray cameras (source-detector projection model) is totally different from usual cameras, and the classical pinhole model is not a good description for these cameras. Thus, it is hard to apply typical hand-eye calibration methods to such cameras with different projection models.

In this paper, we propose a novel approach for hand-eye calibration based on reprojection error minimization. It simultaneously estimates the hand-eye transformation as well as the calibration pattern pose such that the reprojection error of the pattern is minimized. In contrast to traditional methods, it directly takes calibration pattern images, but not camera poses with respect to the pattern. This approach brings several benefits to us. First, it does not require the camera pose for each image explicitly, and thus, it does not rely on PnP (Perspective-n-Points) algorithms which could be unreliable and unavailable for unusual cameras. Second, this method can be applied to various camera models by changing the projection model. The proposed method is implemented as a pose graph optimization problem so that it can be easily extended for new projection models and can estimate the transforms efficiently and robustly. The implementation is available as open source software, from a public repository[3].

The rest of this paper is organized as follows. Sec. II reviews related work on hand-eye calibration. Sec. III describes the proposed hand-eye calibration method and its extension to a source-detector projection model. In Sec. IV, the proposed method is evaluated in simulations with pinhole and source-detector camera models. Sec. V presents evaluations on datasets acquired with real systems. Sec. VI concludes the paper.

II. RELATED WORK

In the past, a number of methods for hand-eye calibration have been proposed. Shiu and Ahmad proposed a homogeneous equation to estimate hand-eye transformation from the camera and reference frame motions [3]. Tsai and Lenz proposed another method which first estimates the translation and then the rotation of the hand-eye transformation [4], and it has been widely used in a lot of systems [5]. A method proposed by Chou and Kamel solves the problem in the same way with quaternion representation which simplifies the estimation problem [6]. Daniilidis proposed a method to simultaneously estimate translation and rotation using dual

Manuscript received: September, 10, 2018; Revised December, 13, 2018; Accepted December, 31, 2018.

This paper was recommended for publication by Editor Dezhen Song upon evaluation of the Associate Editor and Reviewers' comments. *This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 779431.

[1] Kenji Koide and Emanuele Menegatti are with the Department of Information Engineering, University of Padova, via Gradenigo 6/B, 35131, Padova, Italy, koide@dei.unipd.it

Digital Object Identifier (DOI): see top of this page.

[2] http://spirit-h2020.eu/

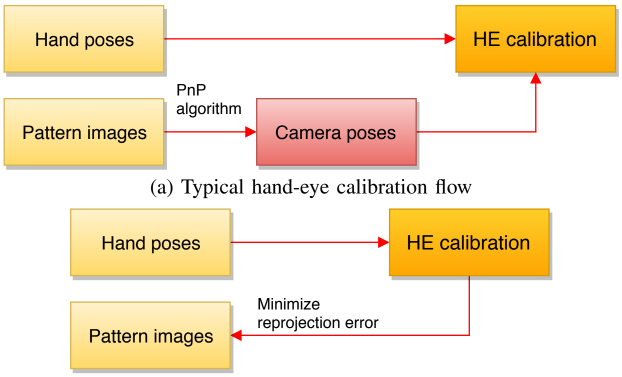
[3] https://github.com/koide3/st_handeye_graph.git

quaternions parameterization and singular value decomposition [7]. Strobl and Hirzinger introduced a new metric for automatic optimal weighting of the translation and the rotation error metrics [8]. Several works proposed automatic hand-eye calibration frameworks [9], [10], in which a robot automatically collects images of the calibration pattern and performs hand-eye calibration, and they have been successfully applied to real systems [11].

Those methods take a set of pairs of robot hand and camera pose and optimize the hand-eye transformation estimate in the pose space. On the one hand, they can be applied to any sensor which can observe the sensor pose with respect to a reference frame, such as LiDAR and IMU. On the other hand, in the case with a camera, we need to explicitly estimate the camera pose for each input image using, for instance, a PnP (Perspective-n-Points) algorithm [12]. Therefore, typical hand-eye calibration flow with a camera is to first collect a set of pairs of hand pose and calibration pattern image (like chessboard), then estimate the camera pose for each image using a PnP algorithm, and finally input the hand poses and the estimated camera poses to a calibration algorithm (see Fig. 1 (a)). The accuracy of those calibration methods, thus, rely on the camera pose estimation accuracy. Although Strobls and Hirzinger discussed that the main source of perturbation is the hand pose error [8], the camera pose estimation error would be another error source. Indeed, similar tasks which require accurate visual information, like visual odometry, suffer from imaging conditions [13]. In our case, for instance, industrial thermographic cameras often have a very narrow field view, and we observed the difficulty to perform PnP algorithms with such cameras in Thermobot[4], a thermography-based inspection robot project. Furthermore, the optics of X-ray cameras is totally different from usual cameras, and the existing methods cannot be applied to such cameras.

The concept of the proposed method is similar to the one proposed by Tabb and Yousef [14]. Their method also estimates the hand-eye transformation such that the reprojection error is minimized. However, their method does not take hand pose errors into account and would suffer from such errors. On contrary, our method properly models hand pose errors, and robust estimation allows us to deal with outliers of the visual and hand pose observations. In addition to that, we propose an extension for hand-eye calibration with an X-ray camera in this work.

X-ray cameras have a projection mechanism so-called source-detector projection. In this projection model, a ray emitted from a *source point* passes on a 3D point and then arrives at a *detector plane* (see Fig. 3). In medical X-ray systems, the *source* and *detector* are mounted on the endpoints of a C-shaped rigid frame (so-called C-arm) so that their relative pose is fixed. In this setting, the X-ray cameras can be modeled as a pinhole camera by introducing a *virtual detector plane* [15], and usual hand-eye calibration techniques can be applied to them [16], [17]. However, in industrial situations, the *source* and *detector* are mounted on separated robot arms so that the X-ray camera can observe a large object, and the existing methods cannot deal with such systems.



(a) Typical hand-eye calibration flow

(b) Proposed hand-eye calibration flow

Fig. 1: While the typical flow requires to estimate the camera pose for each input image, the proposed flow directly takes the input images and minimizes the reprojection error to estimate the hand-eye (HE) transformation.

III. METHODOLOGY

A. Hand-Eye Calibration based on Reprojection Error Minimization

For general and accurate hand-eye calibration, we propose a reprojection error minimization-based calibration method. Reprojection error minimization itself is a well-known technique. It has been applied to various tasks, such as structure from motion [18] and object pose estimation [12]. It has been proven that it outperforms traditional methods in terms of accuracy and robustness. However, only a few works applied this approach to the context of hand-eye calibration [14].

The proposed method directly takes robot hand poses and calibration pattern images and optimizes the hand-eye transformation and the pattern pose such that the reprojection error of the pattern is minimized (see Fig. 1 (b)). It is implemented as a pose graph optimization problem. Pose graph optimization is one of the most successful approaches for pose sequence estimation, such as 2D and 3D SLAM [19], visual SLAM [20], and sensor network calibration [21]. It is able to solve a large non-linear optimization problem efficiently and robustly by utilizing the sparsity of the graph structure.

B. Hand-Eye Calibration with a Pinhole Camera Model

Fig. 2 shows the proposed graph structure for hand-eye calibration with a pinhole camera model. Circles in the graph are vertices which hold parameters to be estimated. In this graph, the hand-eye transformation ($\widehat{HE}$, Hand-to-Eye), the object pose ($\widehat{OW}$, Object-to-World), and the transformation between the pattern and the camera ($\widehat{OE}_i$, Object-to-Eye) are estimated from observations of N hand poses ($\widetilde{WH}_i$, World-to-Hand) and N images of the calibration pattern which consists of M points ($\widetilde{P}_i^j$). Rectangles are edges which connect multiple vertices and calculate the errors between them, and octagons are fixed parameters. A is the camera matrix for

[4] http://thermobot.eu

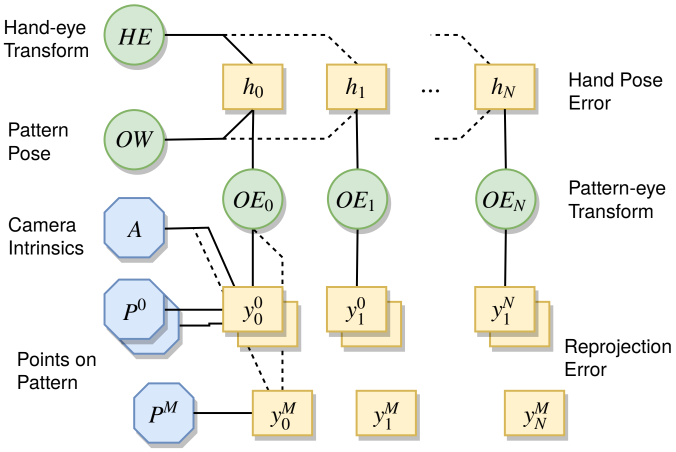


Fig. 2: A graph model for the reprojection error minimization-based hand-eye calibration with a pinhole camera model. Circles are vertices which hold parameters to be estimated. Rectangles are edges which connect vertices and calculate the errors between them, and octagons are fixed parameters.

projection, and P^j is the 3D coordinate of the j-th point on the calibration pattern.

Edge h_i defines the error between an estimated $\widehat{OE}_i$ and the corresponding hand pose observation $\widetilde{WH}_i$:

$$measurement(h_i) = \widetilde{WH}_i, \tag{1}$$

$$error(h_i) = m(\widehat{OW} \cdot \widehat{OE}_i^{-1} \cdot \widehat{HE} \cdot \widetilde{WH}_i), \tag{2}$$

where, function m converts a transformation matrix to the manifold representation, which consists of the translation vector and the vector part of the quaternion without the scalar part $[x, y, z, qx, qy, qz]$, to prevent the corruption of the Gauss-Newton optimization due to the over-parameterization of the rotation [22].

Edge y_i^j projects a 3D point P^j of the calibration pattern into the i-th image space with the estimated $\widehat{OE}_i$ and calculates the reprojection error between the projected point and the observed 2D point $\widetilde{P}_i^j$:

$$measurement(y_i^j) = \widetilde{P}_i^j, \tag{3}$$

$$[x, y, z]^T = A \cdot \widehat{OE}_i \cdot P^j, \tag{4}$$

$$error(y_i^j) = \widetilde{P}_i^j - \frac{1}{z}[x, y]^T. \tag{5}$$

The result of the iterative estimation depends on the initial guess of the hand-eye transformation. For pinhole cameras, we first use Tsai's method [4] to obtain an initial guess, and then apply the graph optimization method.

C. HandEye Calibration with a Source-Detector Camera Model

Here, we extend the graph model to deal with source-detector projection models (see Fig. 3). In this projection model, a ray is emitted from the *source point* s, and it passes on a 3D point v and arrives at the *detector plane* d. Let us assume that s and v are transformed into the detector frame. The projected point (u, v) is given by:

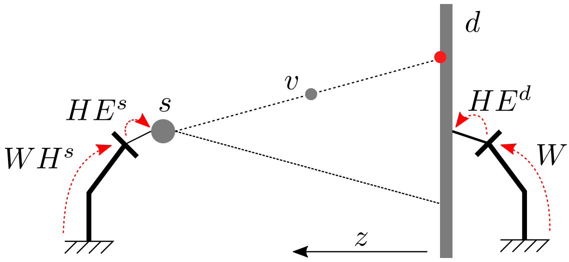


Fig. 3: Source-Detector Projection Model.

$$\begin{bmatrix} x' \\ y' \\ 0 \end{bmatrix} = s + \frac{s.z}{s.z - v.z}(v - s) \tag{6}$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x \cdot x' + c_x \\ f_y \cdot y' + c_y \end{bmatrix}, \tag{7}$$

where f_x, f_y are the inverse of the sensor size of each pixel, and c_x, c_y are the image principal point of the detector.

In this model, we have two robot hands. One is equipped with the detector H^d, and the source is mounted on the other hand H^s. While the transformation between the detector and its hand is represented as a 6DoF transformation $\widehat{HE}^d$, the transformation between the source and its hand is represented as a 3D translation vector $\widehat{HE}^s$ since the source is modeled as a point source in this model.

With this projection model, the graph model is extended as shown in Fig. 4. The parameters to be estimated and the observations are summarized in Table I. While the pattern-detector transformation part (vertices $\widehat{OE}_i^d$ and above) is identical to the graph model for pinhole cameras, new vertices representing the hand-source transformation $\widehat{HE}^s$ and the source position in the world frame $\widehat{S}_i$ are added. Edge q_i calculates the error between the estimated $\widehat{S}_i$ and the source's hand pose $\widetilde{WH}_i^s$:

$$measurement(q_i) = \widetilde{WH}_i^s \tag{8}$$

$$error(q_i) = (\widetilde{WH}_i^s)^{-1}(-\widehat{HE}^s) - \widehat{S}_i. \tag{9}$$

Then, edge y_i^j minimizes the reprojection error:

$$measurement(y_i^j) = \widetilde{P}_i^j, \tag{10}$$

$$\boldsymbol{v}_i^{j\prime} = \widehat{OE}_i^d \cdot P^j, \tag{11}$$

$$\boldsymbol{s}_i' = \widehat{OE}_i^d \cdot \widehat{OW}^{-1} \cdot \widehat{S}_i, \tag{12}$$

$$error(y_i^j) = \widetilde{P}_i^j - Proj^{SD}(\boldsymbol{s}_i', \boldsymbol{v}_i^{j\prime}), \tag{13}$$

where $Proj^{SD}$ is the source-detector projection function given by equation 7.

The graph models are implemented by utilizing g2o, a general hyper graph optimization library [23]. To make the estimation robust to hand pose and visual detection noises, Huber robust kernel [24] is applied to each edge. The graph optimization is performed by the Levenberg-Marquardt solver [25].

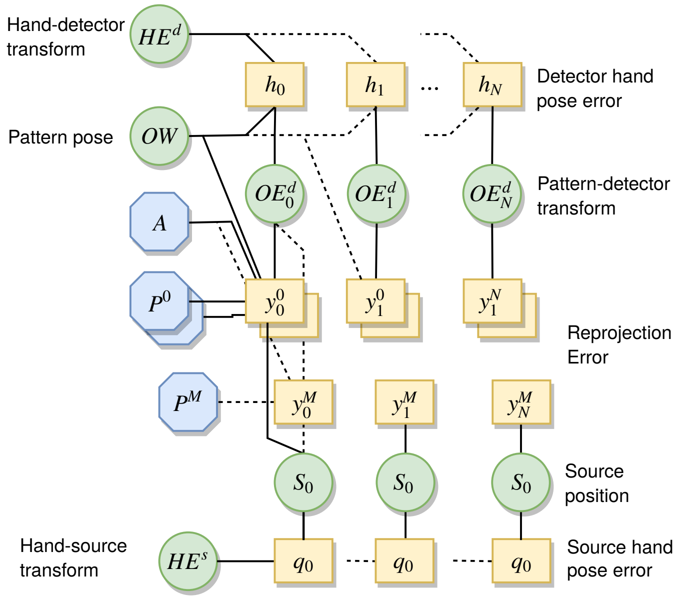


Fig. 4: A graph model for reprojection error minimization-based hand-eye calibration with a source-detector camera model.

TABLE I: Parameters to be estimated and observations in the graph model for source-detector cameras

	Symbol	Meaning
parameters	$\widehat{HE}^d$	hand-detector transformation
	$\widehat{HE}^s$	hand-source translation
	$\widehat{OW}$	pattern-world transformation
	$\widehat{OE}^d$	pattern-detector transformation
	$\widehat{S}_i$	source position in the i-th detector frame
observes	$\widetilde{WH}_i^d$	i-th detector hand pose (edge h_i)
	$\widetilde{WH}_i^s$	i-th source hand pose (edge q_i)
	$\widetilde{P}_i^j$	j-th point observation in the i-th image (edge y_i^j)

IV. Evaluation

A. Simulation with a Pinhole Camera Model

We conducted simulations to evaluate the estimation accuracy of the proposed method under visual detection noise and robot hand pose noise.

For each simulation, we randomly sample a hand-eye transform from the pose distribution where the translation $\|\boldsymbol{t}\| \in \mathcal{N}(\mu = 0, \sigma = 0.3)[m]$, and the rotation angle $\theta \in [0, 90]$ [deg]. We generate nine camera stations at two altitudes with respect to the calibration pattern consisting of (7x5) points, and the camera takes an image at each station while staring at the pattern (see Fig. 5 and 6). Thus, 18 images are generated for each simulation.

The proposed method, Tsai's method [4], the dual quaternions-based method [7], and two nonlinear methods proposed by Tabb and Yousef [14] (*c2* and *rp1*) are applied to estimate the hand-eye transformation. *c2* in [14] estimates the hand-eye transformation which describes the observed hand poses and pattern poses well (so called $AX = ZB$ equation), while *rp1* is based on reprojection error minimization.

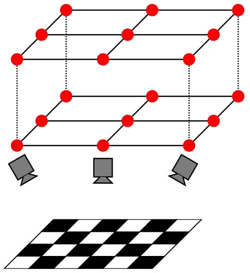


Fig. 5: Simulation setting. 18 images are taken in total for calibration.

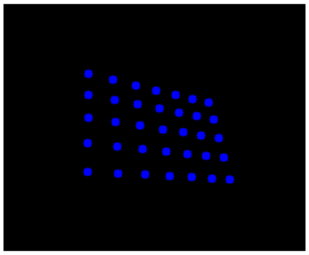


Fig. 6: An example of the simulated calibration pattern images.

While the proposed method and *rp1* directly receive the images, for the other methods, a PnP algorithm [12] is applied to the images to estimate the camera pose with respect to the pattern, and then the estimated camera poses are passed to the methods. While changing the standard deviation of noises on the projected 2D points and the robot hand poses, the simulation is performed 100 times for each noise setting.

Fig. 7 shows the evaluation results. The proposed method and *rp1* show comparable results under visual noise. However, *rp1* shows lower estimation accuracy under hand pose noise, since it does not take hand pose errors into account. *c2* shows better results than Tsai's and dual quaternions-based methods thanks to the non-linear optimization. However, under translation noise, *c2* and Tsai's methods show close accuracy since the noise spans over the linear space, and the nonlinear optimization does not make a difference in such a case. The proposed method shows good estimation results under visual noise and hand pose noise.

B. Simulation with a Source-Detector Camera Model

We conducted simulations with a source-detector camera model as well. Similar to Sec. IV-A, nine camera stations are generated for each of the source and the detector. Images of the calibration pattern are taken while 1) either the source or the detector moves, 2) both of them move to diagonally opposite stations. Thus, the total number of images for each simulation is 27. The detector focuses on the source at every station.

The hand-detector transformation is randomly sampled from the pose distribution where the translation $\|\boldsymbol{t}^d\| \in \mathcal{N}(0, 0.3)$ [m], and the rotation angle $\theta^d \in [0, 15]$ [deg]. The hand-source translation $\boldsymbol{t}^s$ is sampled from $\mathcal{N}(0.0, 0.3)$ [m]. Since we do not have any method to provide an initial guess for the source-detector projection model, we set a smaller maximum rotation angle to the hand-detector distribution to avoid that the estimation converges to a local solution. However, in practical applications, a rough initial guess is often available from the design of the camera fixture. Thus, we consider that the method can be applied to real tasks if it can work under initial rotation errors in this range.

Fig. 8 shows the evaluation result. Under visual noise and hand pose noise, the proposed method converges well, and the maximum translation, and rotation errors are lower than 0.1 [m] and 2.0 [deg], respectively. While the rotation error was kept very low through the evaluation, the translation error of

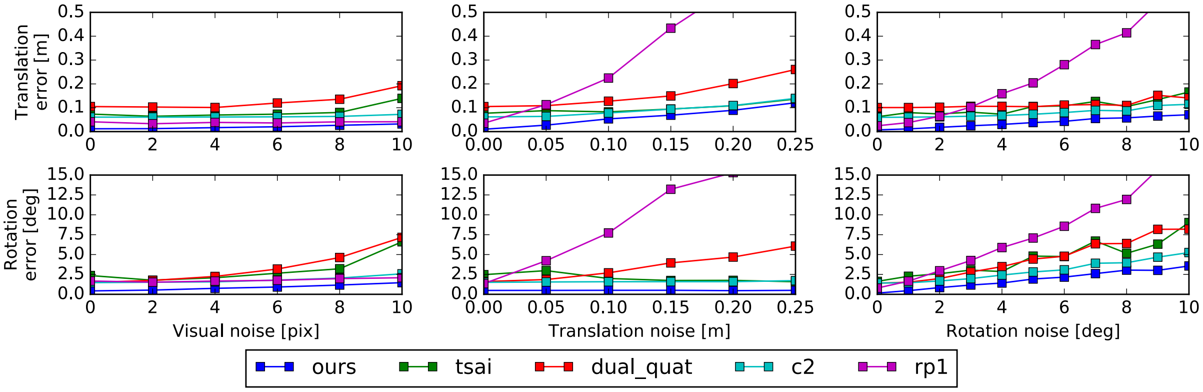


Fig. 7: Hand-eye transformation estimation evaluation with a pinhole camera model.

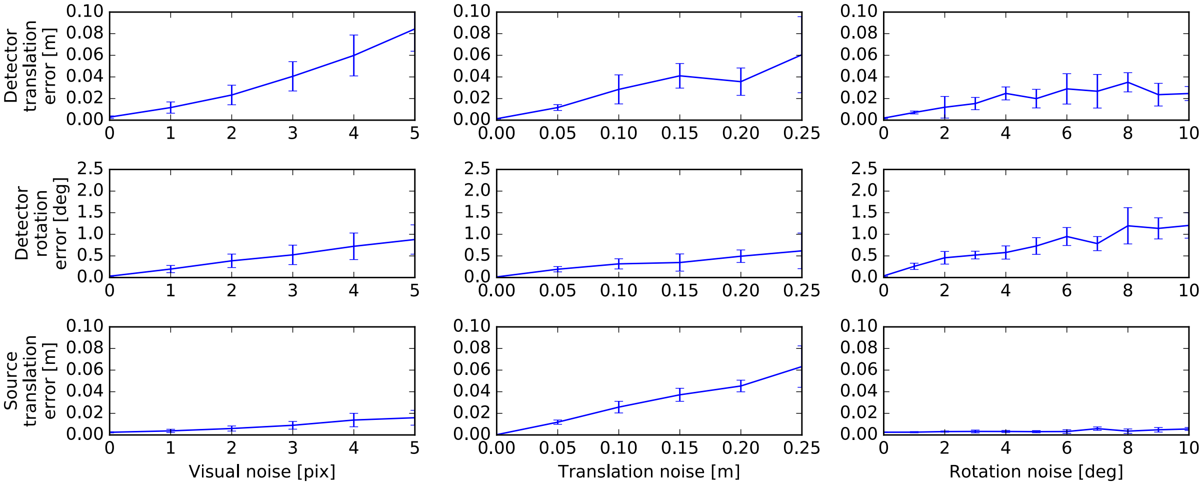


Fig. 8: Hand-eye transformation estimation evaluation with a source-detector camera model. The error bar indicates the standard deviation.

the hand-detector transformation got larger under large visual and hand pose translation noises. In particular, visual noise has a significant impact on the detector translation estimation. It is worth to mention that, the accuracy of the detector translation estimation can be improved by increasing the number of camera stations and the camera station interval. In the case we take two images at each station while slightly changing the view direction (10 [deg]), the detector translation error decreases from 83.4 [mm] to 71.7 [mm]. The error gets further decreased to 51.8 [mm] by increasing the station interval from 0.5 [m] to 0.75[m].

V. Experiment

A. Hand-eye Calibration with a Pinhole Camera

We applied the proposed method to two datasets acquired with real robot hands and sensors.

The first one is a public dataset provided by [14]. It consists of a set of datasets, and we used four of them which contain 88, 28, 36, and 20 pairs of calibration pattern image and hand pose, respectively. In this dataset, two metrics, RRMSE (Reprojection Root Mean Square Error) and RAE (Reconstruction Accuracy Error), are exploited to evaluate calibration methods. RRMSE is defined by:

$$RRMSE = \sqrt{\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\|\widetilde{P}_i^j - \widehat{P}_i^j\|^2}, \qquad (14)$$

where $\widehat{P}_i^j$ is the j-th point reprojected onto the i-th image. RAE is defined by:

$$RAE = \frac{1}{M}\sum_{j}^{M}\|\widehat{P}^j - \widehat{OW}\cdot P^j\|^2, \qquad (15)$$

where $\widehat{P}^j$ is the 3D coordinate of each point on the calibration pattern. In our case, we estimate $\widehat{P}^j$ such that the reprojection error is minimized with Nelder-Mead method [26].

Table II shows the evaluation results. As explained in Sec IV-A, *c2* and *rp1* [14] are based on $AX = ZB$ equation and reprojection error minimization, respectively. *rp2* is a variation of *rp1*, in which the camera intrinsic parameters are also optimized by reprojection error minimization. *rp2* shows the lowest reprojection errors on all the datasets. However, it often shows worse reconstruction accuracy. This result suggests that it overfitted to the reprojection error term due

TABLE II: Reprojection error and 3D reconstruction error evaluation on the dataset [14]

	Dataset1			Dataset2			Dataset3			Dataset4		
Method	Time	RRMSE	RAE	Time	RRMSE	RAE	Time	RRMSE	RAE	Time	RRMSE	RAE
c2 [14]	**0.511**	1.67244	5.82014	**0.199**	3.96036	0.08633	**0.256**	3.55362	1.92736	**0.142**	2.38787	1.71460
rp1 [14]	14.548	1.56905	0.23821	5.789	3.74123	**0.02950**	10.706	2.89018	1.45425	6.454	1.60321	1.43699
rp2 [14]	63.688	**1.36292**	0.52353	66.702	**3.18114**	0.44990	46.056	**2.67963**	1.58750	18.414	**1.50483**	1.48145
ours	1.000	1.56160	**0.21371**	0.334	3.75133	0.03760	0.496	2.87851	**1.33743**	0.272	1.61900	**1.37951**

Time [sec], RRMSE [pix], RAE [mm^2]

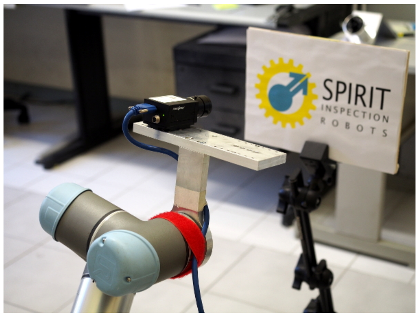


Fig. 9: A snapshot of the experiment in a real environment.

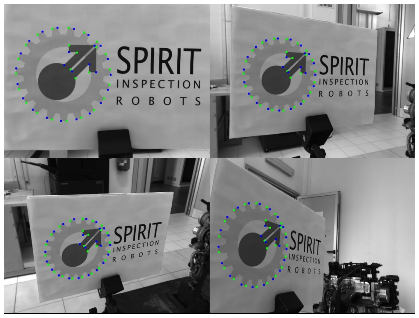


Fig. 10: The images for reconstruction and the keypoint correspondences.

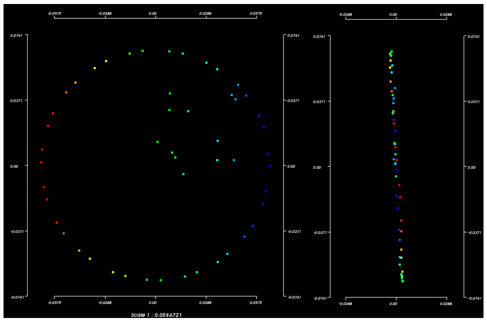


Fig. 11: Reconstructed 3D point cloud. The color indicates the "X"-coordinate of each point.

to the intrinsic parameter optimization. The proposed method and *rp1* show better reprojection accuracy than *c2* thanks to the reprojection error minimization approach. In terms of the reconstruction accuracy, the proposed method shows the best results except for the dataset 2. However, the reconstruction error on this dataset is much smaller than the other datasets, and the difference between the errors of our method and *rp1* is very small (0.008 [mm^2]). Although our method requires a processing time which is almost double with respect to *c2*, it takes only one second even on the dataset 1 (with 88 images), and it is much faster than the other reprojection error-based methods thanks to the efficient graph optimization. Note that the processing time of our method includes Tsai's method for initial guess estimation.

We also took a dataset in our environment to validate the calibration methods in a practical situation. We used a robot hand (Universal Robots, UR5) equipped with a monocular camera (PointGrey, Flea3). For calibration, we took 18 images of a standard asymmetric calibration pattern (4x11) in the same way as described in Sec IV-A. The hand-eye transformation was estimated from those images using the proposed method, Tsai's method [4], dual quaternions-based method [7], and *c2* and *rp1* in [14].

Then, for 3D reconstruction, we took four images of a pattern (the project logo) with random camera poses. We made a set of keypoints on the pattern and corresponding points on each image by hand (see Fig. 10). For each keypoint, we estimate the 3D position such that the reprojection error is minimized. Since the 3D coordinates of the keypoints are unknown in this dataset, we calculate 3D geometrical flatness error instead of RAE. The flatness error is defined as the mean absolute error between the reconstructed points and the fitted plane.

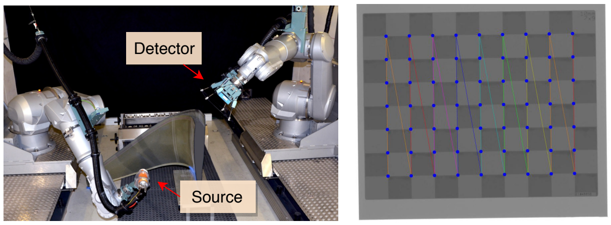
TABLE III: 3D reconstruction error evaluation

Error	ours	Tsai [4]	DQ [7]	c2 [14]	rp1 [14]
RRMSE [pix]	**7.118**	11.901	11.488	7.229	11.245
3D flatness [m]	**0.0203**	0.0216	0.0214	0.0208	0.0221

Fig. 11 shows the 3D point cloud reconstructed with the hand-eye transformation estimated by the proposed method. The shape of the pattern is well captured, and the points mostly lie on a flat plane correctly. Table III shows the error metrics of the reconstructed points with the hand-eye transformations estimated by the different calibration methods. The reconstructed points with the proposed method have smaller errors in both the metrics. Since the error sources of the reconstruction other than the hand-eye transformation (hand poses and camera parameters) were fixed for all the calibration methods, this result suggests that the proposed method provides a better hand-eye transform estimation than the existing methods. *rp1* shows worse results than *c2* in this case. We consider that this is caused by the hand pose error due to oscillations.

B. Hand-eye Calibration with an X-ray Camera

We conducted a hand-eye calibration experiment using a real X-ray imaging system. Fig. 12 (a) shows a snapshot of the imaging system equipped with two arms with the X-ray source and detector. We made a chessboard calibration pattern with PCB (printed circuit board), and Fig. 12 (b) shows an example X-ray image of the calibration pattern. This system has been calibrated based on the CAD model, and it does not provide the hand poses, but provides only the source and detector poses. Since the hand-eye transformations and the robot hand poses of this system were not available to us, we assume virtual hand-eye transformations, and add the virtual transformations to the provided source and detector poses.



(a) X-ray imaging system (b) X-ray image

Fig. 12: A snapshot of the X-ray imaging system and an example X-ray image of the calibration pattern. Blue dots indicate the corners of the checkerboard reprojected with the estimated hand-eye transformations.

We set $\|\boldsymbol{t}^s\| = 0.3\text{m}, \|\boldsymbol{t}^d\| = 0.3[\text{m}], \theta^d = 30[\text{deg}]$, and the translation and rotation directions are randomly sampled. Then, we estimate the virtual transformations from the x-ray images and the generated virtual hand poses to validate the proposed method. We took 34 images in total while moving and rotating the source and the detector.

The graph optimization method successfully converged, and the translation and rotation errors of the detector were 0.196 [mm] and 0.0143 [deg], the source translation error was 1.29 [mm]. Blue dots in Fig. 12 (b) indicate the corners of the checkerboard reprojected with the estimated transforms. We found that, in the X-ray system, there is no nuisance factor which causes distortion (like lens of usual cameras), and the optics is well described by the source-detector projection model. This result shows the applicability of the proposed method to real X-ray imaging systems.

VI. Conclusion and Future Work

This paper presented a hand-eye calibration method based on reprojection error minimization. It can deal with different camera models by changing the projection model, and the graph optimization scheme allows us to estimate the hand-eye transformation efficiently and robustly. Through the simulations and the experiments, it has been confirmed that the proposed method provides better estimation than existing methods, and it can be applied to real systems.

In addition to hand-eye calibration, hand-eye synchronization also plays an important role in continuous scan tasks. To realize a "general" hand-eye synchronization method, we plan to extend the reprojection error minimization-based method to the temporal space.

Acknowledgment

This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 779431.

We would like to thank S. Ayalur-Karunakaran from FACC AG, Ried im Innkreis, Austria for acquiring and providing the X-ray images.

References

[1] M. Antonello, M. Munaro, and E. Menegatti, "Efficient measurement of fibre orientation for mapping carbon fibre parts with a robotic system," in *Intelligent Autonomous Systems 14*. Springer, 2017, pp. 757–769.

[2] N. Correll, K. E. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J. M. Romano, and P. R. Wurman, "Analysis and observations from the first amazon picking challenge," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 1, pp. 172–188, 2018.

[3] Y. C. Shiu and S. Ahmad, "Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form ax=xb," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 1, pp. 16–29, 1989.

[4] R. Tsai and R. Lenz, "Real time versatile robotics hand/eye calibration using 3d machine vision," in *IEEE International Conference on Robotics and Automation*. IEEE, 1988, pp. 554–561 vol.1.

[5] E. Marchand, F. Spindler, and F. Chaumette, "Visp for visual servoing: a generic software platform with a wide class of robot control skills," *IEEE Robotics & Automation Magazine*, vol. 12, no. 4, pp. 40–52, 2005.

[6] J. C. K. Chou and M. Kamel, "Finding the position and orientation of a sensor on a robot manipulator using quaternions," *The International Journal of Robotics Research*, vol. 10, no. 3, pp. 240–254, 1991.

[7] K. Daniilidis, "Hand-eye calibration using dual quaternions," *The International Journal of Robotics Research*, vol. 18, no. 3, pp. 286–298, 1999.

[8] K. Strobl and G. Hirzinger, "Optimal hand-eye calibration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006, pp. 4647–4653.

[9] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 345–358, 1989.

[10] M. Antonello, A. Gobbi, S. Michieletto, S. Ghidoni, and E. Menegatti, "A fully automatic hand-eye calibration system," in *European Conference on Mobile Robots*. IEEE, 2017, pp. 1–6.

[11] M. Munaro, M. Antonello, M. Moro, C. Ferrari, G. Clemente, E. Pagello, and E. Menegatti, "FibreMap: Automatic Mapping of Fibre Orientation for Draping of Carbon Fibre Parts," in *Workshop on ROS-Industrial in European Research Projects*, 2014, pp. 272–275.

[12] J. A. Hesch and S. I. Roumeliotis, "A direct least-squares (dls) method for pnp," in *IEEE International Conference on Computer Vision*. IEEE, 2011, pp. 383–390.

[13] M. O. A. Aqel, M. H. Marhaban, M. I. Saripan, and N. B. Ismail, "Review of visual odometry: types, approaches, challenges, and applications," *Springer*, vol. 5, no. 1, p. 1897, 2016.

[14] A. Tabb and K. M. A. Yousef, "Solving the robot-world hand-eye(s) calibration problem with iterative methods," *Machine Vision and Applications*, vol. 28, no. 5-6, pp. 569–590, 2017.

[15] N. Navab, M. Mitschke, and O. Schtz, "Camera-augmented mobile c-arm (CAMC) application: 3d reconstruction using a low-cost mobile c-arm," in *Medical Image Computing and Computer-Assisted Intervention*. Springer, 1999, pp. 688–697.

[16] M. Mitschke and N. Navab, "Recovering the x-ray projection geometry for three-dimensional tomographic reconstruction with additional sensors: Attached camera versus external navigation system," *Medical Image Analysis*, vol. 7, no. 1, pp. 65–78, 2003.

[17] R. R. Galigekere, K. Wiesent, and D. W. Holdsworth, "Cone-beam reprojection using projection-matrices," *IEEE Transactions on Medical Imaging*, vol. 22, no. 10, pp. 1202–1214, 2003.

[18] O. Edlund, "A software package for sparse orthogonal factorization and updating," *ACM Transactions on Mathematical Software*, vol. 28, no. 4, pp. 448–482, 2002.

[19] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2d lidar slam," in *IEEE International Conference on Robotics and Automation*. IEEE, 2016, pp. 1271–1278.

[20] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: A versatile and accurate monocular slam system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[21] M. Munaro, F. Basso, and E. Menegatti, "Openptrack: Open source multi-camera calibration and people tracking for rgb-d camera networks," *Robotics and Autonomous Systems*, vol. 75, pp. 525–538, 2016.

[22] C. Hertzberg, "A framework for sparse, non-linear least squares problems on manifolds," 2008.

[23] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G: A general framework for graph optimization," in *IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 3607–3613.

[24] P. J. Huber, "Robust estimation of a location parameter," in *Springer Series in Statistics*. Springer, 1992, pp. 492–518.

[25] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of Applied Mathematics*, vol. 2, no. 2, pp. 164–168, 1944.

[26] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.